普通高等教育“十三五”规划教材

最优化理论与方法

Optimization Theory and Methods

王景恒 编著

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

普通高等教育“十三五”规划教材

最优化理论与方法

王景恒　编著

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

内 容 简 介

本书是为应用数学专业本科生、工科硕士研究生所编写的一门最优化课程教材，是作者综合多年的教学实践，在原有教学讲义基础上，经过反复修订而成的。

本书主要内容包括线性规划及其对偶理论、最优性条件、无约束最优化问题和约束最优化问题。每部分内容都较全面系统地介绍了其基本理论和优化算法。作为教材，每章后附有习题，以便加深学生对所学知识的理解和掌握。

本书除作为教材外，也可作为从事最优化方面工作的科研人员和工程技术人员的学习参考用书。

图书在版编目（CIP）数据

最优化理论与方法 / 王景恒编著. —北京：北京理工大学出版社，2018.8
ISBN 978-7-5682-6276-7

Ⅰ. ①最…　Ⅱ. ①王…　Ⅲ. ①最佳化理论②最优化算法　Ⅳ. ①O224②O242.23

中国版本图书馆 CIP 数据核字（2018）第 203595 号

出版发行 / 北京理工大学出版社有限责任公司
社　　址 / 北京市海淀区中关村南大街 5 号
邮　　编 / 100081
电　　话 /（010）68914775（总编室）
（010）82562903（教材售后服务热线）
（010）68948351（其他图书服务热线）
网　　址 / http://www.bitpress.com.cn
经　　销 / 全国各地新华书店
印　　刷 / 保定市中画美凯印刷有限公司
开　　本 / 710 毫米×1000 毫米　1/16
印　　张 / 20.25
字　　数 / 300 千字
版　　次 / 2018 年 8 月第 1 版　2018 年 8 月第 1 次印刷
定　　价 / 46.00 元

责任编辑 / 封　雪
文案编辑 / 封　雪
责任校对 / 周瑞红
责任印制 / 王美丽

图书出现印装质量问题，请拨打售后服务热线，本社负责调换

前　言

最优化是用数学方法研究最优方案的选择问题，是数学的一个重要分支，它在诸多领域有着十分广泛的应用，最优化问题的基本知识已成为新的工程技术和管理人员所必备的基础知识. 因此，最优化理论与方法是目前高等院校普遍开设的一门数学课程.

本书是为应用数学专业本科生、工科硕士研究生所写的一门最优化课程教材，是笔者结合多年的教学实践，在原有教学讲义的基础上，经过多次的教学实践修改而成的. 本书较全面、系统地介绍了最优化的基本理论和方法. 本书在阐述问题时力求清晰、透彻，语言浅显易懂、深入浅出，并试图让学生了解算法的来龙去脉，以便使他们在解决实际问题的过程中更好地运用这些方法.

本书共有 11 章. 第 1 章绪论，向读者介绍最优化作为学科的发展过程以及最优化基本概念和基本问题. 为了便于读者更好地学习本书知识，特增加两节数学预备知识. 第 2 章线性规划和第 3 章线性规划对偶问题，主要涉及线性规划的基本内容. 第 4 章最优性条件，讨论最优化问题解的必要与充分条件. 第 5 章算法，主要介绍求解最优化问题迭代算法的基本问题. 第 6 章一维搜索，第 7 章使用导数的最优化方法和第 8 章无约束最优化的直接方法，主要讨论无约束最优化问题的求解方法. 第 9 章二次规划，第 10 章可行方向法和第 11 章惩罚函数法，主要讨论约束最优化问题的求解方法. 作为教材，在每章的后面均列有习题，便于学生复习和巩固该章所学的知识.

本书在编排上采用分块式，即线性规划、无约束最优化问题和约束最优化问题三大块，各块相对独立，教师可根据授课内容和对象进行适当删减或增加.

本书除作为教材外，也可作为从事最优化方面工作的科研人员和工程技术人员的学习参考用书.

本书在编写和出版过程中，同行专家提出了许多宝贵意见和建议，同时也得到了长春理工大学教材建设项目的资助，以及北京理工大学出版社的大力支持，在此一并表示衷心感谢.

限于编者的水平，书中不妥与错误之处在所难免，殷切期望专家、学者不吝指教.

编　者

2018 年 3 月

目　录

第 1 章 绪 论

1.1 引言

最优化就是针对给出的实际问题，从众多的可行方案中选出最优方案．其任务是讨论研究决策问题的最佳选择之特性，构造寻求最佳解的计算方法．最优化问题广泛见于工程设计、经济规划、生产管理、交通运输、国防等重要领域．例如，在工程设计中，怎样选择设计参数，使得设计方案既能满足设计需求，又能降低成本；在资源分配中，怎样分配有限资源，使得分配方案既能满足各方面的基本要求，又能获得好的经济效益；在生产计划安排中，选择怎样的计划方案才能提高产值和利润；在城建规划中，怎样安排布局才能有利于城市发展；在运输问题中，满足相应要求的条件下，选择怎样的路径才能使运输线路最短；在军事指挥中，如何制订作战方案，使之能有效地消灭敌人，保存自己，有利于战争的全局，等等．在人类活动的各个领域中，此类问题，不胜枚举．最优化为这些问题的解决提供了理论基础和求解方法．

最优化既是一个古老的课题，又是一门年轻的学科．早在 17 世纪，英国科学家牛顿创立微积分的时代，就已提出极值问题，后来拉格朗日研究一个函数在一组等式约束条件下的极值问题时提出了乘数法．1847 年，法国数学家柯西研究了函数值沿什么方向下降最快的问题，提出了最速下降法．1939 年，苏联数学家 Л. В. Канторович 提出了解决下料问题和运输问题这两种线性规划的求解方法．人们关于最优化问题的研究工作，随着历史的发展不断深入．但是，任何科学的进步，都受到历史条件的限制，直到 20 世纪 40 年代，最优化这个古老课题还尚未形成独立的有系统的学科．

20 世纪 40 年代以来，生产和科学研究突飞猛进地发展，特别是电子计算机日益广泛应用，使最优化问题的研究不仅成为一种迫切需要，而且有了求解的有力工具，最优化理论和算法才得以迅速发展，并不断完善，逐步成为一门系统的学科．1947 年，美国科学家 Dantzig 提出了求解线性规划问题的单纯形法，为线性规划的理论和算法奠定了基础．1951 年，由 Kuhn 和 Tucker 完成了非线性规划的理论基础性工作．20 世纪 50 年代以后，人们从一些自然现象和规律中受到启发，提出了许多求解复杂优化问题的新方法．例如，1953 年，Metropolis 最早提出了模拟退火算法；1975 年，Holland 教授在他的专著中比较系统地论述了遗传算法；1992 年，Dorigo 在他的博士论文中首先提出了一种全新的蚁群系统启发式算法，在此基础上蚁群算法逐渐发展起来．总之，近半个世纪以来，最优化方法得到了充分研究，在理论上取得了非常重要的研究成果，在实际应用中正在发挥越来越大的作用．最优化方法已经成为发展迅速、内容丰富、应用广泛的活跃的学科．

1.2 最优化问题

1.2.1 最优化问题数学模型

最优化问题通常可以表示为数学规划问题．数学规划是指对含有 n 个变量的目标函数求极值，而这些变量也可能受到某些条件（等式方程或不等式方程）的限制，其一般数学表达式为

$$\min f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbf{R}^n$$
$$\text{s.t.}\begin{cases} c_i(\boldsymbol{x}) = 0, & i \in E = \{1,2,\cdots,l\} \\ c_i(\boldsymbol{x}) \geqslant 0, & i \in I = \{l+1,\cdots,l+m\} \end{cases} \tag{1.2.1}$$

其中，$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称为决策变量；$f(\boldsymbol{x})$ 称为目标函数；$c_i(\boldsymbol{x}) = 0(i \in E)$ 和 $c_i(\boldsymbol{x}) \geqslant 0(i \in I)$ 称为约束条件；min 和 s.t. 分别是英文单词 minmum（极小化）和 subject to（受限于）的缩写．

根据实际问题的不同要求，最优化模型有不同的形式，但经过适当的变换都可以转换成上述一般形式．例如，若求 $\max f(\boldsymbol{x})$，可以将目标函数写成 $\min(-f(\boldsymbol{x}))$，若不等式约束为 $c_i(\boldsymbol{x}) \leqslant 0$，则可以写成 $-c_i(\boldsymbol{x}) \geqslant 0$．

在问题（1.2.1）中，若 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})$ $(i \in E \cup I)$ 均为线性函数，则相应的规划称为线性规划；若 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})$ $(i \in E \cup I)$ 中含有非线性函数，则称为非线性规划．

问题（1.2.1）也称为约束最优化问题．若去掉问题（1.2.1）的约束条件，则得到

$$\min f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbf{R}^n \tag{1.2.2}$$

称为无约束最优化问题．

下面给出最优化问题解的概念．

定义 1.2.1 对于约束最优化问题（1.2.1），满足约束条件的点称为可行点，全体可行点组成的集合称为可行域，记作 D，即

$$D = \{\boldsymbol{x} \mid c_i(\boldsymbol{x}) = 0, i \in E;\ \ c_i(\boldsymbol{x}) \geqslant 0,\ \ i \in I,\ \ \boldsymbol{x} \in \mathbf{R}^n\}$$

无约束最优化问题（1.2.2）的可行域为整个空间．

定义 1.2.2 设 $f(\boldsymbol{x})$ 为目标函数，D 为可行域，$\bar{\boldsymbol{x}} \in D$．若对每个 $\boldsymbol{x} \in D$，$f(\boldsymbol{x}) \geqslant f(\bar{\boldsymbol{x}})$ $(f(\boldsymbol{x}) > f(\bar{\boldsymbol{x}}))$ 成立，则称 $\bar{\boldsymbol{x}}$ 为 $f(\boldsymbol{x})$ 在 D 上的（严格）全局极小点．

定义 1.2.3 设 $f(\boldsymbol{x})$ 为目标函数，D 为可行域．若存在 $\bar{\boldsymbol{x}} \in D$ 的 $\varepsilon > 0$ 邻域 $N(\bar{\boldsymbol{x}}, \varepsilon) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \bar{\boldsymbol{x}}\| < \varepsilon\}$，使得对每个 $\boldsymbol{x} \in D \cap N(\bar{\boldsymbol{x}}, \varepsilon)$，$f(\boldsymbol{x}) \geqslant f(\bar{\boldsymbol{x}})$ $(f(\boldsymbol{x}) > f(\bar{\boldsymbol{x}}))$ 成立，则称 $\bar{\boldsymbol{x}}$ 为 $f(\boldsymbol{x})$ 在 D 上的一个（严格）局部极小点．

对于极大化问题，可以类似地定义全局极大点和局部极大点．

根据上述定义，全局极小点也是局部极小点，而局部极小点不一定是全局极小点．但是对于某些特殊情形，如将在后面介绍的凸规划，局部极小点也是全局极小点．

1.2.2 最优化问题举例

1．无约束最优化问题

例 1.2.1 曲线拟合问题

设有两个物理量 η 和 ξ，根据某一物理定律得知，它们满足如下关系：

$$\eta = a + b\xi^c$$

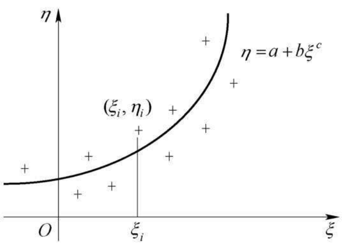

图 1.2.1 曲线拟合问题

其中 a,b,c 三个常数在不同情况下取不同的值. 现由实验得到一组数据 $(\xi_1,\eta_1),(\xi_2,\eta_2),\cdots,(\xi_m,\eta_m)$.

试选择 a,b,c 的值，使曲线 $\eta = a + b\xi^c$ 尽可能靠近所有的实验点 (ξ_i,η_i) , $i=1,2,\cdots,m$ ，如图 1.2.1 所示.

解析 这个问题可用最小二乘原理求解，即选择 a,b,c 的一组值，使得偏差的平方和

$$\delta(a,b,c) = \sum_{i=1}^{m}(a + b\xi_i^c - \eta_i)^2 \tag{1.2.3}$$

达到最小. 换句话说，就是求 3 个变量的函数 $\delta(a,b,c)$ 的极小点作为问题的解.

为了便于今后的讨论，我们把它写成统一的形式，把 a,b,c 换成 x_1,x_2,x_3 ，记为 $\boldsymbol{x} = (x_1,x_2,x_3)^{\mathrm{T}}$ ，把 δ 换成 f ，这样问题就归结为求解

$$\min f(\boldsymbol{x}) = \sum_{i=1}^{m}(x_1 + x_2\xi_i^{x_3} - \eta_i)^2 \tag{1.2.4}$$

2. 约束最优化问题

例 1.2.2 生产计划问题

设某工厂用 4 种资源生产 3 种产品，每单位第 j 种产品需要第 i 种资源的数量为 a_{ij} ，可获得利润为 c_j ，第 i 种资源总消耗量不能超过 b_i ，试问如何安排生产才能使总利润最大？

解析 设 3 种产品的产量分别为 x_1，x_2，x_3 ，这是决策变量，目标函数是总利润 $c_1x_1 + c_2x_2 + c_3x_3$ ，约束条件是资源限制 $a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 \leqslant b_i (i=1,2,3,4)$ 和产量非负限制 $x_j \geqslant 0 (j=1,2,3)$. 问题概括为，在一组约束条件下，确定一个最优生产方案 $\boldsymbol{x}^* = (x_1^*,x_2^*,x_3^*)^{\mathrm{T}}$ ，使目标函数值最大，数学模型如下:

$$\max f(\boldsymbol{x}) = \sum_{j=1}^{3} c_j x_j$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{3} a_{ij}x_j \leqslant b_i, & i=1,2,3,4 \\ x_j \geqslant 0, & j=1,2,3 \end{cases}$$

例 1.2.3　选址问题

设有 n 个市场，第 j 个市场的位置为 (a_j, b_j)，对某种货物的需要量为 $q_j(j=1,2,\cdots,n)$. 现计划建立 m 个货栈，第 i 个货栈的容量为 $c_i(i=1,2,\cdots,m)$. 试确定货栈的位置，使各货栈到各市场的运输量与路程乘积之和最小.

解析　设第 i 个货栈的位置为 $(x_i, y_i)(i=1,2,\cdots,m)$，第 i 个货栈供给第 j 个市场的货物量为 $w_{ij}(i=1,\cdots,m; j=1,\cdots,n)$，第 i 个货栈到第 j 个市场的距离为 d_{ij}，一般定义为

$$d_{ij}=\sqrt{(x_i-a_j)^2+(y_i-b_j)^2} \tag{1.2.5}$$

或

$$d_{ij}=|x_i-a_j|+|y_i-b_j| \tag{1.2.6}$$

我们的目标是使运输量与路程乘积之和最小，如果按式（1.2.5）定义，就是使

$$\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2}$$

最小. 约束条件是：

（1）每个货栈向各市场提供的货物量之和不能超过它的容量；

（2）每个市场从各货栈得到的货物量之和应等于它的需要量；

（3）运输量不能为负数.

因此，问题的数学模型如下：

$$\min f(\boldsymbol{x},\ \boldsymbol{y},\ \boldsymbol{w})=\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2}$$

$$\text{s.t.}\begin{cases}\displaystyle\sum_{j=1}^{n} w_{ij}\leqslant c_i, & i=1,\cdots,m\\ \displaystyle\sum_{i=1}^{m} w_{ij}=q_j, & j=1,\cdots,n\\ w_{ij}\geqslant 0, & i=1,\cdots,m;\quad j=1,\cdots,n\end{cases}$$

在上述例子中，例 1.2.1 是无约束最优化问题. 例 1.2.2 和例 1.2.3 是约束最优化问题. 同时例 1.2.2 模型中，目标函数和约束函数都是线性的，因此又属于线性规划，而例 1.2.1 和例 1.2.3 模型中含有非线性函数，因此又属于非线性规划.

1.3 数学基础

1.3.1 向量与矩阵

1．向量

设 $\boldsymbol{x}$， $\boldsymbol{y}$ 为 n 维欧几里得（Euclid）空间中的两个向量，即

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$$
$$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$$

定义向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积为

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\sum_{i=1}^{n}x_i y_i=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \tag{1.3.1}$$

由式（1.3.1），容易得到内积具有如下性质：

（1）$\langle\boldsymbol{x},\boldsymbol{x}\rangle\geqslant 0$，且 $\langle\boldsymbol{x},\boldsymbol{x}\rangle=0$ 的充要条件是 $\boldsymbol{x}=\mathbf{0}$；

（2）$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{y},\boldsymbol{x}\rangle$；

（3）$\langle\lambda\boldsymbol{x}+\mu\boldsymbol{y},\boldsymbol{z}\rangle=\lambda\langle\boldsymbol{x},\boldsymbol{z}\rangle+\mu\langle\boldsymbol{y},\boldsymbol{z}\rangle$.

常用的向量范数有 L_1 范数， L_2 范数和 L_∞ 范数，分别为

$$\|\boldsymbol{x}\|_1=\sum_{j=1}^{n}|x_j| \tag{1.3.2}$$

$$\|\boldsymbol{x}\|_2=(\sum_{j=1}^{n}x_j^2)^{\frac{1}{2}} \tag{1.3.3}$$

$$\|\boldsymbol{x}\|_\infty=\max_j|x_j| \tag{1.3.4}$$

一般地，对于 $1\leqslant p<\infty$， L_p 范数为

$$\|\boldsymbol{x}\|_p=(\sum_{j=1}^{n}|x_j|^p)^{\frac{1}{p}} \tag{1.3.5}$$

这里应指出，上述向量范数中，$\|\boldsymbol{x}\|_2$ 称为欧几里得范数．如无特殊指明，后面将用 $\mathbf{R}^n$ 表示 n 维欧几里得空间.

范数具有如下性质：

（1）$\|\boldsymbol{x}\| \geqslant 0$，且 $\|\boldsymbol{x}\| = 0$ 的充要条件是 $\boldsymbol{x} = 0$；

（2）$\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$；

（3）$\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$（三角不等式）；

（4）$|\langle \boldsymbol{x},\ \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|$（柯西-施瓦茨不等式）.

且等式成立的充要条件是：$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 共线，即存在 λ，使得

$$\boldsymbol{x} = \lambda \boldsymbol{y}$$

将柯西-施瓦茨不等式写成分量形式为

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leqslant \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2 \right)^{\frac{1}{2}} \tag{1.3.6}$$

2．矩阵

设 $m \times n$ 阶矩阵为

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}_{m \times n}$$

把矩阵 $\boldsymbol{A}$ 的行（或列）的极大线性无关组的向量个数称为矩阵 $\boldsymbol{A}$ 的秩，记为 $\mathrm{rank}(\boldsymbol{A})$.

若

$$\mathrm{rank}(\boldsymbol{A}) = \min\{m, n\} \tag{1.3.7}$$

则称矩阵 $\boldsymbol{A}$ 是满秩的. 若 $m<n$，则称为行满秩. 若 $m>n$，则称为列满秩. 当 $m = n$ 时，则矩阵为 n 阶非奇异（可逆）方阵.

当 $\boldsymbol{A}$ 为方阵时，计算矩阵 $\boldsymbol{A}$ 的行列式为

$$\det(\boldsymbol{A}) = \sum_{i_1 i_2 \cdots i_n} (-1)^{\tau(i_1 i_2 \cdots i_n)} a_{1 i_1} a_{2 i_2} \cdots a_{n i_n} \tag{1.3.8}$$

其中 $\tau(i_1 i_2 \cdots i_n)$ 是 $i_1 i_2 \cdots i_n$ 的逆序数，$\sum\limits_{i_1 i_2 \cdots i_n}$ 表示对所有的 n 阶排列求和.

矩阵非奇异的充要条件是：

$$\det(\boldsymbol{A}) = 0 \tag{1.3.9}$$

设 $\boldsymbol{A}$ 为 $n \times n$ 阶矩阵，若 $\boldsymbol{A}$ 满足

$$A^{\mathrm{T}} = A \tag{1.3.10}$$

则称 A 为对称矩阵；若对于一切 $x \neq 0$，均有

$$x^{\mathrm{T}} A x > 0 \tag{1.3.11}$$

则称 A 为正定矩阵；若对于一切 x，均有

$$x^{\mathrm{T}} A x \geqslant 0 \tag{1.3.12}$$

则称 A 为半正定矩阵.

1.3.2 序列的极限

定义 1.3.1 设 $\{x^{(k)}\}$ 是 $\mathbf{R}^n$ 中一个向量序列，$\bar{x} \in \mathbf{R}^n$，如果对每个任给的 $\varepsilon > 0$，存在正整数 k_ε，使得当 $k > k_\varepsilon$ 时，有 $\|x^{(k)} - \bar{x}\| < \varepsilon$，则称序列收敛到 $\bar{x}$，记作 $\lim\limits_{k\to\infty} x^{(k)} = \bar{x}$.

按此定义，序列若存在极限，则任何子序列有相同的极限，即序列的极限是唯一的.

定义 1.3.2 设 $\{x^{(k)}\}$ 是 $\mathbf{R}^n$ 中一个向量序列，如果存在一个子序列 $\{x^{(k_j)}\}$，使 $\lim\limits_{k_j\to\infty} x^{(k_j)} = \hat{x}$，则称 $\hat{x}$ 是序列 $\{x^{(k)}\}$ 的一个聚点.

根据定义易知，如果无穷序列有界，即存在正数 M，使得对所有 k 均有 $\|x^{(k)}\| \leqslant M$，则这个序列必有聚点.

定义 1.3.3 设 $\{x^{(k)}\}$ 是 $\mathbf{R}^n$ 中一个向量序列，如果对任意给定的 $\varepsilon > 0$，总存在正整数 k_ε，使得当 m，$l > k_\varepsilon$ 时，有 $\|x^{(m)} - x^{(l)}\| < \varepsilon$，则 $\{x^{(k)}\}$ 称为柯西序列.

在 $\mathbf{R}^n$ 中，柯西序列有极限.

定理 1.3.1 设 $\{x^{(j)}\} \subset \mathbf{R}^n$ 为柯西序列，则 $\{x^{(j)}\}$ 的聚点必为极限点.（证明从略）

1.3.3 多变量函数的梯度、Hesse 矩阵和泰勒展开式

定义 1.3.4 设 $f(x)$ 为多变量函数，则称 n 维列向量

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^{\mathrm{T}} \tag{1.3.13}$$

为函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的梯度.

定义 1.3.5　设 $f(\boldsymbol{x})$ 为多变量函数，则称 $n\times n$ 矩阵 $\nabla^2 f(\boldsymbol{x})$，其中第 i 行第 j 列元素

$$[\nabla^2 f(\boldsymbol{x})]_{ij}=\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j},\ 1\leqslant i,\ j\leqslant n \tag{1.3.14}$$

为函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的 Hesse 矩阵.

当 $f(\boldsymbol{x})$ 为二次函数时，梯度及 Hesse 矩阵很容易求得. 二次函数可以写成下列形式：

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x}+\boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}+\delta$$

式中，$\boldsymbol{G}$ 是 n 阶对称矩阵，$\boldsymbol{r}$ 是 n 维列向量，δ 是常数. 函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的梯度

$$\nabla f(\boldsymbol{x})=\boldsymbol{G}\boldsymbol{x}+\boldsymbol{r}$$

Hesse 矩阵

$$\nabla^2 f(\boldsymbol{x})=\boldsymbol{G}$$

下面给出方向导数概念.

定义 1.3.6　对于任意给定的 $\boldsymbol{d}\neq\boldsymbol{0}$，若极限

$$\lim_{\alpha\to 0^+}\frac{f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\alpha\|\boldsymbol{d}\|}$$

存在，则该极限值为 $f(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处沿方向 $\boldsymbol{d}$ 的一阶方向导数，记为 $\dfrac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}})$，即

$$\frac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}})=\lim_{\alpha\to 0^+}\frac{f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-f(\bar{\boldsymbol{x}})}{\alpha\|\boldsymbol{d}\|} \tag{1.3.15}$$

由定义求方向导数是很困难的，这里给出它的另一种表达式.

定理 1.3.2　若函数 $f(\boldsymbol{x})$ 具有连续的一阶偏导数，则它在 $\bar{\boldsymbol{x}}$ 处沿方向 $\boldsymbol{d}$ 的一阶方向导数为

$$\frac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}})=\left\langle \nabla f(\bar{\boldsymbol{x}}),\frac{\boldsymbol{d}}{\|\boldsymbol{d}\|}\right\rangle=\frac{1}{\|\boldsymbol{d}\|}\boldsymbol{d}^{\mathrm{T}}\nabla f(\bar{\boldsymbol{x}}) \tag{1.3.16}$$

证明　记 $\bar{\boldsymbol{x}}=(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n)^{\mathrm{T}}$，$\boldsymbol{d}=(d_1,\ d_2,\cdots,\ d_n)^{\mathrm{T}}$，考虑单变量函数

$$\varphi(\alpha)=f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})=f(\bar{x}_1+\alpha d_1,\ \bar{x}_2+\alpha d_2,\cdots,\ \bar{x}_n+\alpha d_n) \tag{1.3.17}$$